

## Summary

### Summary

Methodological News contains articles on a range of topics across different areas within the Methodology Division in the ABS. This issue contains 3 articles on Protecting the Confidentiality of Providers, Measuring Statistical Impacts with State Space Models and Big Data and Data Integration.

## Protecting the Confidentiality of Providers

### PROTECTING THE CONFIDENTIALITY OF PROVIDERS

The release of aggregated statistics provides useful information for researchers, policy decision-makers and the community, however, by definition comes with a risk of privacy loss - known as the risk-utility trade-off. The ABS publishes data in a way that maximises the utility of statistical information while upholding our obligations to protect the confidentiality of our providers. The ABS applies a suite of rigorous confidentiality methods and protections in order to be able to release useful information while making it not likely to enable identification of individuals.

The ABS undertakes a range of investigations to stay at the leading edge of methodological practice in confidentiality - particularly in light of increasingly sophisticated computing technologies and algorithms available to attackers, and the emerging types of data and information solutions being expected by analysts and the community.

Differential Privacy is a framework for protecting privacy by adding random noise to released data. This framework specifically focuses on providing 'plausible deniability' to a respondent around whether they can be identified as contributing information to that dataset. This is expressed through a measurable limit to the resulting loss of privacy. Differential Privacy promises a number of theoretical benefits, such as measuring the privacy loss associated with using a particular perturbation distribution and also calculating the accumulated privacy loss across multiple queries. However, it must also be considered in terms of underlying assumptions, practical implementation, social licence and stakeholder acceptance, and the broader confidentiality protections / legislative requirements. In particular:

1. The privacy loss measure relates only to the perturbation component, so will underestimate the protection provided by the overall package of confidentiality applied to the data under the 'Five Safes' framework;
2. There are limitations when extending beyond person-level counting type queries;
3. Access to correlated auxiliary data weakens the protection provided by (any) perturbation mechanism; and
4. Implementation requirements may not perturb data beyond levels acceptable to analysts or be practically infeasible.

It is important to note that the theory is still evolving, however, there may be aspects that are potentially useful to augment current approaches. The ABS is assessing the potential benefits of differential privacy and identifying unresolved issues and implications, including implications for practical implementation. The ABS is also engaging with experts and other National Statistical Organisations who face similar challenges to better understand these implications. The US Census Bureau, for example, are incorporating perturbation as part of their confidentialisation method for the 2020 US Census, including an adapted form of differential privacy; and are working to identify and resolve outstanding issues and progress the theoretical foundations.

It is important to remember that the differential privacy framework does not guarantee privacy protection, rather it allows a bound to be placed on the maximum privacy loss for aggregated tables – thereby providing a data-providing agency with a mechanism to control risk. Differential privacy also does not in itself inform the analyst of the utility-loss relevant to their particular analysis - some literature argued that under Differential Privacy, the usefulness of microdata files may be severely damaged.

While there is a growing literature investigating potential relaxations, adaptations and alternatives that attempt to address the issues, the theory is still maturing. Some of the issues include aggregated units (eg, families), descriptive statistics, weighted data, magnitude data, correlated data and longitudinal data. In particular, there is

currently no widely accepted way to calculate the accumulated privacy loss over multiple queries (current composition strategies are based on a 'worst case' scenario). The ABS is undertaking work to assess the differential privacy measure relevant to the perturbation mechanisms commonly applied.

For more information, please contact Daniel Elazar [Methodology@abs.gov.au](mailto:Methodology@abs.gov.au)

The [ABS Privacy Policy](#) outlines how the ABS will handle any personal information that you provide to us.

## Measuring Statistical Impacts with State Space Models

### MEASURING STATISTICAL IMPACTS WITH STATE SPACE MODELS

The ABS is embarking on a transformation program, which includes a change in ABS systems (editing, imputation, estimation, etc.), use of different collection modes and list sampling frames etc. Once this transformation is completed, it is expected to deliver positive changes to the production of official statistics. However, there is also a risk of introducing impacts on some ABS time series. The ABS is working on developing methods to measure, and where needed, to adjust for any impacts on such time series.

A number of different State Space models were considered in this study, including the difference model for parallel collection, phase-in model with and without Kalman filter initialisation using estimated impacts from parallel collection and the full model that utilises both parallel collection and phase-in period information. The models' performance is evaluated from the results of a simulation study for the Labour Force Survey (LFS). The full model with parallel collection performs the best in terms of the power of detecting overall impacts when compared to the other models.

For more information, please contact Oksana Honchar [Methodology@abs.gov.au](mailto:Methodology@abs.gov.au)

The [ABS Privacy Policy](#) outlines how the ABS will handle any personal information that you provide to us.

## Big Data and Data Integration

### BIG DATA AND DATA INTEGRATION

Efficiency and effectiveness considerations, as well as government and community expectations, require the ABS to consider new ways of ingesting Big Data in the production of official statistics. Big Data, also known as 'found data', due to the very nature of their creation, generally suffer from coverage bias and measurement errors. Unless corrected, Big Data cannot be used for producing high quality official statistics often used in informing important decisions.

In this research, we considered the practical implementation of a Regression Data Integration (RDI) method, developed jointly by the ABS and a researcher from the Iowa State University, to correct for this bias using existing ABS estimation processes. The study demonstrates that the RDI method can be readily applied within ABS statistical production processes to produce estimates of total and associated sampling errors. The results show that the method has potential to achieve significant gains where little measurement error correction is needed in either dataset. When measurement error needs to be taken into account, the gains may be eroded.

For more information, contact Lyndon Ang [Methodology@abs.gov.au](mailto:Methodology@abs.gov.au)

The [ABS Privacy Policy](#) outlines how the ABS will handle any personal information that you provide to us.

## How to Contact Us and Email Subscriber List

### HOW TO CONTACT US AND EMAIL SUBSCRIBER LIST

Methodological News features articles and developments in relation to methodology work done within the ABS Methodology Division. By its nature, the work of the division brings it into contact with virtually every other area of the ABS. Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

If you would like to be added to or removed from our electronic mailing list, please contact:

Nick Husek  
Methodology Division  
Australian Bureau of Statistics  
Locked Bag No. 10  
Belconnen ACT 2617

Email: [Methodology@abs.gov.au](mailto:Methodology@abs.gov.au)

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.

## About this Release

Methodological News is a quarterly information bulletin from the Methodology Division. The newsletter features articles and developments in relation to work done with the division.

---

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence